ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 70 – 2003

# Consistent estimation of dynamic panel data models with time-varying individual effects

**Céline NAUGES, Alban THOMAS**[1]

**ABSTRACT**. – This paper proposes a new specification for dynamic panel data models, where unobserved heterogeneity is modelled by the sum of the usual additive individual effect, and a multiplicative, time-varying individual effect. We show that usual GMM estimators based on first-difference or quasi-difference transformations are generally inconsistent with our model specification, and we propose a consistent GMM estimation procedure based on a double-difference transformation. Small-sample properties of alternative GMM estimators are investigated through Monte Carlo experiments.

---

## Estimation convergente de modèles dynamiques de panel avec effets individuels variant dans le temps

**RÉSUMÉ.** – Cet article présente un nouveau modèle dynamique de panel dans lequel le terme d'erreur admet deux effets individuels, l'un sous forme additive et l'autre associé à un paramètre variant dans le temps. Nous montrons que les estimateurs GMM usuels ne sont pas convergents en général dans ce cas, et nous proposons un estimateur GMM convergent basé sur une double différenciation (différence-première et quasi-différence), dont les propriétés sont étudiées à l'aide de simulations de Monte-Carlo.

[1] Corresponding address: LEERNA-INRA, Université des Sciences Sociales de Toulouse, 21 Allée de Brienne, 31000 Toulouse, France
Emails: cnauges@toulouse.inra.fr, thomas@toulouse.inra.fr.
Phone +33 (0) 5 61 12 85 21, Fax +33 (0) 5 61 12 85 20.

# 1 Introduction

In the usual dynamic panel data model, it is well known that the fixed-effect estimator is biased when the number of periods is fixed (see HSIAO [1986]). Instrumental-Variable procedures have been advocated as a convenient way of dealing with the endogeneity bias in the model in first difference, originating from the correlation between the lagged dependent variable or additional regressors and the transformed error term. The literature on dynamic panel data has been recently unified to some extent (AHN and SCHMIDT [1995], [1997]), as GMM (Generalized Method of Moments) has now become the most popular way of estimating such models. GMM estimation typically proceeds by considering a set of orthogonality conditions for the model in first difference (to eliminate the individual effect), and building an instrument matrix from dependent variables lagged two periods or more, see ARELLANO and BOND [1991], AHN and SCHMIDT [1995, 1997], BLUNDELL and BOND [1998].

Another specification less often encountered in the panel data literature is when unobserved heterogeneity may vary across time periods, *i.e.*, is affected by a time-varying parameter. This corresponds to a multiplicative two-way model, in which time effects are treated as non-random parameters. Possible applications include the effect of a macro economic regulation policy on a sample of firms, whose unobserved productivity parameters are modified after regulation takes place; technological change; modification of unobserved consumption patterns through time. KIEFER [1980] derives a Concentrated Least Squares estimator for this model, and HOLTZ-EAKIN, NEWEY and ROSEN [1988] deal with the case of time-varying individual effects in the framework of vector autoregressive models. The model may be estimated by a quasi-differencing procedure and GMM estimation can be used, by considering orthogonality conditions in which time effects appear jointly with parameters of interest (see AHN, LEE and SCHMIDT [2001]).

A drawback when using the above model in empirical applications is that no individual effect appears separately in the error term. In practice, such a specification forbids cases where a component only of the unobserved heterogeneity is affected by time effects. It then seems natural to consider a mixed specification with both time-varying and time-independent individual effects.[2] Such a specification contains the two usual models described above as special cases, and allows the econometrician to model unobserved heterogeneity as the sum of a time-invariant and a time-varying component.

Let us introduce some examples where such a specification could be useful. Consider first a production function for firm $i$ at time $t$, that relates output level $Q_{it}$ to labor and capital inputs ($L_{it}$ and $K_{it}$), and incorporates a time-varying component to account for technological change, $A_t = h(t)^{\gamma}$. For a Cobb-Douglas technology we would have $Q_{it} = A_t L_{it}^{\alpha_1} K_{it}^{\alpha_2} e^{u_{it}}$, where $u_{it}$ is an error term with mean 0. Assume that the technological change component

2. Multiplicative individual effects in the variance for dynamic panel data have been investigated by MEGHIR and WINDMEIJER [1999].

is individual-specific: $A_t = h(t)^{\gamma_i}$, where $\gamma_i$ is a random coefficient, and consider the usual error component structure for panel data: $u_{it} = \eta_i + \varepsilon_{it}$. The above equation becomes

$$\log Q_{it} = \gamma_i \log h(t) + \alpha_1 \log L_{it} + \alpha_2 \log K_{it} + \eta_i + \varepsilon_{it}.$$

The output supply equation now admits two random, individual-components $\gamma_i$ and $\eta_i$. This model generalizes the usual specification for the technological change component in production functions, which would be $\gamma_i = \gamma$ and $h(t) = t$.

Another example is based on LILLARD and WEISS [1979], who specify a wage equation in which worker's ability is decomposed into an individual-specific component $\eta_i$ (before the worker is hired) and a time-varying component capturing on-the-job increases in ability: $\eta_i + \alpha_i(t - \bar{t})$, where $\bar{t}$ denotes the date of hiring. These authors assume that on-the-job training raises worker's ability linearly, and a more general specification would obtain by considering an increasing function $h(t - \bar{t})$ instead of $(t - \bar{t})$.

A final example is based on the $q$ investment model, whose simplest form is

$$\left(\frac{I}{K}\right)_{it} = a + b\left(\frac{I}{K}\right)_{i,t-1} + cq_{it} + u_{it}, \tag{1}$$

where $I_{it}$ is investment expenditures of firm $i$ during period $t$, $K_{it}$ is the replacement value of the capital stock, and $q_{it}$ is Tobin's $q$ of firm $i$ (market value of invested goods over their replacement cost). Heterogeneity in firm investment responses has often been advocated as a major reason for many unsatisfactory estimation results concerning this equation. HSIAO, PESARAN and TAHMISCIOGLU [1999] estimate the model above by allowing parameters $a$, $b$ and $c$ to be heterogenous across firms. The applied researcher may wish, moreover, to include a trend component that would capture firms' sensitivity to overall market conditions. If firm responses are also heterogenous with respect to underlying macroeconomic variables, the model would then be

$$\left(\frac{I}{K}\right)_{it} = a_i + b\left(\frac{I}{K}\right)_{i,t-1} + cq_{it} + \theta_t\alpha_i + u_{it}, \tag{2}$$

where $\theta_t$ is the trend component and $\alpha_i$ the associated firm-specific slope parameter. In the model specification above, intercept $a_i$ and slope $\alpha_i$ are two distinct firm effects.

Once assuming a mixed specification with both time-varying and time-independent individual effects, an important issue is the consistency of GMM estimators based on usual (first-difference or quasi-difference) transformations, when instruments include lagged dependent variables, and the mixed structure is the true data generating process. First-difference filtering would eliminate the usual, additive individual effect, while quasi difference would filter out the multiplicative one, but eliminating both individual effects would *a priori* require an alternative model transformation.

In this paper we propose a GMM estimation procedure for models with mixed covariance structure, that exploits moment restrictions in the same way as for the usual dynamic panel data models. The key difference is the model transformation we suggest to filter out individual effects from the original error term. Our GMM estimator is consistent under alternative data generating processes (additive or multiplicative individual effects). As in the multiplicative individual effect case suggested by HOLTZ-EAKIN, NEWEY and ROSEN [1988], orthogonality conditions are non-linear, which requires numerical minimization of the GMM criterion function. This is because active covariance restrictions are written as the product of two functions, one depending on time effects, and the other one depending on parameters of interest only (including, in particular, the autoregressive parameter).

The paper is organized as follows. In section 2, we present the model with a general covariance structure for dynamic panel data. Conditions for equivalence with usual one-way panel data models are discussed. Section 3 deals with GMM estimation in the framework of our mixed covariance structure. An inappropriate choice for the instrument matrix or, equivalently, for the transformation matrix, may lead to inconsistent estimates in the sense that GMM estimators based on first difference or quasi difference and Arellano-Bond or Ahn-Schmidt instruments may not be consistent. We exhibit conditions under which this is the case, and we propose a consistent transformation based on double differencing. Covariance restrictions are then obtained in the same way as AHN and SCHMIDT [1995], allowing us to identify the total number of moment restrictions available. In section 4, we consider an augmented model with explanatory variables, and derive associated moment conditions. Section 5 is devoted to Monte Carlo experiments, in which we assess the properties of GMM estimators in small samples, under alternative assumptions on the data generating process. The mean and standard error of GMM estimators based on three alternative model transformations are computed for different data generating processes, together with associated sizes of the Hansen test. Section 6 concludes by providing guidelines for estimation and testing of dynamic panel data models with time-varying individual effects.

# 2 The model with general covariance structure

Consider the following dynamic model:

(3) $$y_{it} = \delta y_{i,t-1} + u_{it} \qquad i = 1,2,...,N, \quad t = 1,2,...,T,$$

where $\delta$ is the autoregressive parameter (we assume $|\delta| < 1$). Initial values $y_{i0}, i = 1,\dots,N$ are observed and stochastic.[3]

3. The analysis would not change substantially if exogenous explanatory variables were introduced. This question is addressed in section 4.

We are interested in a general form for $u_{it}$ that would contain a purely additive individual effect and a multiplicative error component. We assume that individual heterogeneity is characterized by two individual effects denoted $\alpha_i$ and $v_i$, such that:

$$u_{it} = \alpha_i + \theta_t v_i + \varepsilon_{it}. \tag{4}$$

$\alpha_i$ stands for the usual, linear individual effect, while $\theta_t v_i$ captures an additional, time-varying individual effect. $\alpha_i$ and $v_i$ may not have the same distribution (or second-order moments), and may be correlated. Let $\rho_{\alpha_i v_i}$ denote the correlation coefficient between $\alpha_i$ and $v_i$. The $\theta_t, t = 1,\ldots,T$, are treated as fixed nuisance parameters to be estimated, which modulates the impact of $v_i$ on $y_{it}$. $\varepsilon_{it}$ is the usual econometric error term. As for asymptotics, we consider the most common case: $N \to \infty$ and $T$ is fixed. Throughout the paper, we assume that $\alpha_i$ and $v_i$ have zero mean for all $i$, that $\varepsilon_{it}$ have zero mean for all $i,t$, and that the following assumptions hold:

$$\text{for all } i, \quad cov(\varepsilon_{it}, y_{i0}) = 0 \text{ for all } t, \tag{5}$$

$$\text{for all } i, \quad cov(\varepsilon_{it}, \alpha_i) = 0 \text{ for all } t, \tag{6}$$

$$\text{for all } i, \quad cov(\varepsilon_{it}, v_i) = 0 \text{ for all } t, \tag{7}$$

$$\text{for all } i, \quad cov(\varepsilon_{it}, \varepsilon_{is}) = 0 \text{ for all } t \neq s. \tag{8}$$

Conditions (5) to (8) correspond to the set of assumptions (SA.1)-(SA.3) in AHN and SCHMIDT [1995].[4] Further, we suppose that the observations are independently, but not necessarily identically, distributed across individuals. Variances of $\alpha_i$ and $v_i$ are denoted $\sigma^2_{\alpha_i}$ and $\sigma^2_{v_i}$ respectively, and $\sigma_{\alpha_i v_i}$ represents the covariance between both individual effects.

It is worth noting that two conventional models appear as special cases under the following conditions.

### Condition set 1

Model (4) is equivalent to a Linear One-Way (LOW) model $u_{it} = \mu_i + \varepsilon_{it}$ for the errors if one of the following conditions holds:

$$\theta_t = \bar{\theta} \quad \forall t = 1,2,\ldots,T \tag{9}$$

$$\text{for all } \; i, \; \sigma^2_{v_i} = 0. \tag{10}$$

If condition (9) or condition (10) holds, we trivially obtain a LOW model with $\mu_i = \alpha_i + \bar{\theta} v_i$ or $\mu_i = \alpha_i$ respectively. When any of these conditions holds, both models are equivalent.

4. Identical results obtain using the weaker set of assumptions (MA.1)-(MA.3) in AHN and SCHMIDT [1995], *i.e.*, assuming that covariances in (5)-(8) above are the same for all $t$, rather than zero for all $t$.

**Condition set 2**

Model (4) is equivalent to a Multiplicative Two-Way (MTW) model $u_{it} = m_t \mu_i + \varepsilon_{it}$ for the errors if one of the following conditions holds:

$$(11) \qquad |\rho_{\alpha_i v_i}| = 1 \text{ and } \sigma_{\alpha_i}/\sigma_{v_i} = \bar{\sigma}_{\alpha,v} \quad \forall i;$$

$$(12) \qquad \text{for all } i,\ \sigma^2_{\alpha_i} = 0.$$

When condition (11) holds, the model errors can then be written either as $(\theta_t + \overline{\sigma}_{\alpha,v})v_i + \varepsilon_{it}$ if $\rho_{\alpha_i v_i} = 1$, or $(\theta_t - \overline{\sigma}_{\alpha,v})v_i + \varepsilon_{it}$ if $\rho_{\alpha_i v_i} = -1$. Hence we obtain a MTW model with $m_t = (\theta_t \pm \overline{\sigma}_{\alpha,v})$ and $\mu_i = v_i$. If condition (12) holds we have $\alpha_i = 0$ and thus a MTW model with $m_t = \theta_t$ and $\mu_i = v_i$.

# 3 GMM estimation

The purpose of this section is to produce moment conditions from which a consistent GMM estimator can be constructed, under assumptions (5)-(8) above. We first present usual GMM estimators for the LOW and MTW models and show that associated orthogonality conditions are not satisfied when the true data generating process follows specification (4), unless very specific restrictions are imposed on the parameters. Valid moment conditions based on a double-difference transformation of the general covariance model are then exhibited, and their number is compared to the number of conditions identified from covariance restrictions, in the same spirit as AHN and SCHMIDT [1995].

## 3.1 Usual procedure for the LOW and MTW models

In the LOW model,

$$\begin{aligned} y_{it} &= \delta y_{i,t-1} + u_{it} \\ u_{it} &= \alpha_i + \varepsilon_{it}, \quad i = 1,\ldots,N,\ t = 1,\ldots,T, \end{aligned}$$

under assumptions (5), (6) and (8), the following $T(T-1)/2$ orthogonality conditions built from first differencing the model, are available:

$$(13) \qquad E(y_{is} \triangle u_{it}) = 0 \quad \forall t \geqslant 2, \quad s = 0,\ldots,t-2,$$

where $\triangle$ is the first-difference operator: $\triangle y_{it} = y_{it} - y_{i,t-1}$. See ARELLANO and BOVER [1995], ARELLANO and BOND [1991], AHN and SCHMIDT [1995], [1997] for a discussion of this set of conditions.[5]

5. More recently, BLUNDELL and BOND [1998] proposed the use of an extended linear GMM estimator that uses lagged differences of $y_{it}$ as instruments for equations in levels, in addition to lagged levels of $y_{it}$ as instruments for equations in first differences (see also ARELLANO and BOVER [1995]).

Based on these orthogonality conditions, a consistent GMM estimator of the parameter of interest is obtained by minimizing

$$\left(\frac{1}{N}\sum_{i=1}^{N} W_i'\triangle u_i\right)' A \left(\frac{1}{N}\sum_{i=1}^{N} W_i'\triangle u_i\right),$$

where

$$W_i = \begin{bmatrix} y_{i0} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & y_{i0} & y_{i1} & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & y_{i0} & \cdots & y_{i,T-2} \end{bmatrix},$$

$\triangle u_i = [\triangle u_{i2}, \triangle u_{i3}, \ldots, \triangle u_{iT}]'$ and $A$ is an initial positive definite matrix, and an efficient two-step version of it is obtained using

$$A = \left(\frac{1}{N}\sum_{i=1}^{N} W_i'\widehat{\triangle u_i}\widehat{\triangle u_i}'W_i\right)^{-1},$$

where $\widehat{\triangle u_i}$ are residuals based on an initial consistent estimation of $\delta$.

Assuming assumptions (5), (7) and (8), in the MTW model where $u_{it} = \theta_t v_i + \varepsilon_{it}$, for consistent estimation, the model has to be transformed using a quasi-difference procedure: we substract from the model at time $t$ the lagged equation multiplied by $r_t = \theta_t/\theta_{t-1}$. The corresponding orthogonality conditions, from which, similarly to above, a consistent GMM estimator may be constructed are in this case:[6]

(14) $$E[y_{is}(u_{it} - r_t u_{i,t-1})] = 0 \quad \forall t \geqslant 2, \quad s = 0, \ldots, t-2.$$

Note that, contrary to the LOW model in which a single parameter ($\delta$) has to be estimated, in the case of the MTW model we have $T$ parameters to estimate ($\delta$ and $\{r_t\}, t = 2, \ldots, T$).

It is easy to show that the moment conditions (13) and (14) are not satisfied when considering the mixed error-component model, so that a GMM estimator based on either (13) or (14) will not be consistent. This is because an individual effect (or a function of) remains in the model error after the first-difference or the quasi-difference transformation.

As a matter of fact, consider first model (3)-(4) and assume that conditions (9) and (10) are not satisfied. First-differencing yields the following transformed error:

$$u_{it} - u_{i,t-1} = (\theta_t - \theta_{t-1})v_i + \varepsilon_{it} - \varepsilon_{i,t-1}$$

which still depends on $v_i$. Consider now the model (3)-(4) and assume that conditions (11) and (12) do not hold. If quasi-differencing is applied to the model error, we obtain:

$$u_{it} - r_t u_{i,t-1} = (1 - r_t)\alpha_i + \varepsilon_{it} - r_t\varepsilon_{i,t-1}.$$

6. As the model becomes non-linear in the parameters, a closed-form solution does not exist and a two-step non linear GMM procedure is required in order to obtain consistent estimates of $r_t, t = 1, 2, \ldots, T$ and $\delta$.

Hence, the transformed error is still a function of the individual effect $\alpha_i$ and the resulting GMM estimator will not be consistent, unless $r_t = 1$ for all $t$.

## 3.2 Identifying moment conditions for the general covariance model

We establish the following proposition:

PROPOSITION. *Under assumptions (5)-(8):*
*a) A GMM estimator of model (3)-(4) with instruments including lagged dependent variables and based on the two-step transformation consisting in first-differencing first and then quasi-differencing is consistent.*
*b) A GMM estimator of model (3)-(4) based on quasi-differencing first and then first-differencing is not consistent when instruments include lagged dependent variables, unless $\theta_t = a\theta_{t-1}$ for $a \neq 0$ and for all t.*

This proposition first states that a double-difference procedure allows one to filter out both individual effects: first-differencing eliminates $\alpha_i$ and the transformed error is $v_i \triangle\theta_t + \triangle\varepsilon_{it}$. If we substract from it the differenced equation lagged one time period and multiplied by $\triangle\theta_t/\triangle\theta_{t-1}$ we have:

$$\triangle u_{it} - \tilde{r}_t \triangle u_{i,t-1} = \triangle\varepsilon_{it} - \tilde{r}_t \triangle\varepsilon_{i,t-1},$$

$i = 1,2,\ldots,N$, $t = 3,4,\ldots,T$, where
$\tilde{r}_t = \triangle\theta_t/\triangle\theta_{t-1} = (\theta_t - \theta_{t-1})/(\theta_{t-1} - \theta_{t-2})$.

Result b) in the proposition concerns the permutability of difference operators in our model. When applying quasi-differencing before first-differencing, we obtain the transformed error:

$$\triangle\left[u_{it} - r_t u_{i,t-1}\right] = \triangle\varepsilon_{it} - \triangle\left[r_t \varepsilon_{i,t-1}\right] - \alpha_i \triangle r_t,$$

which still depends on $\alpha_i$, unless $\triangle r_t = 0$. This is the case when $\theta$ is a geometric series with $\theta_t = a\theta_{t-1}$ for all $t$ with $a \neq 0$, because $\triangle r_t$ would then be $\theta_t/\theta_{t-1} - \theta_{t-1}/\theta_{t-2} = a - a = 0$.[7]

The derivation of the moment conditions that are satisfied in the double error-component model is now straightforward. Given assumptions (5)-(8), it is easy to check that the following restrictions must be valid :

$$E\left[y_{is}(\triangle u_{it} - \tilde{r}_t \triangle u_{i,t-1})\right] = 0 \quad t = 3,\ldots,T \quad s = 0,\ldots,t-3.$$

We have $(T-1)(T-2)/2$ such conditions, which are equivalent to those in ARELLANO and BOND [1991]. These moment conditions are non linear in the parameters, as the nuisance parameters $\tilde{r}_t$'s are estimated jointly with the parameter of interest, $\delta$.

7. Interestingly, note that in the special case where $\theta_t$ is an arithmetic series, *i.e.*, $\theta_t = \theta_{t-1} + a$ with $a \neq 0$, the residual transformed by double difference does not depend on parameters $\theta_t$, as $\tilde{r}_t$ would then be $\tilde{r}_t = (\theta_t - \theta_{t-1})/(\theta_{t-1} - \theta_{t-2}) = 1$ for all $t$. In that case, a closed-form expression of GMM estimators using the double-difference transformation can then be obtained.

Second, in a similar fashion to AHN and SCHMIDT [1995] for the LOW case above, we have the $(T-3)$ additional restrictions:

$$E\left[u_{iT}(\triangle u_{it} - \tilde{r}_t \triangle u_{i,t-1})\right] = 0 \qquad t = 3, \ldots, T-1.$$

The total number of these "intuitive" conditions is therefore $(T-1)(T-2)/2 + (T-3)$.

We now investigate additional restrictions that can be found when assuming homoskedasticity of the $\varepsilon_{it}$'s. If the following condition is valid:

$$\text{for all } i, \;\; Var(\varepsilon_{it}) \text{ is the same } \forall t,$$

we have $(T-2)$ restrictions that can be expressed as

$$E\left[\bar{u}_i(\triangle u_{it} - \tilde{r}_t \triangle u_{i,t-1})\right] = 0 \;\; \forall t = 3, \ldots, T,$$

where $\bar{u}_i = 1/T \sum_t u_{it}$. The condition above is similar to the one exhibited by AHN and SCHMIDT [1995]. We do not include in this discussion the moment conditions exhibited by BLUNDELL and BOND [1998], as Ahn and Schmidt conditions are redundant in this context. Moreover, it proves difficult to match the corresponding covariance restrictions with these new moment conditions.

We now wish to verify that the "intuitive" moment conditions identified above maximize the number of conditions for GMM estimation. To see this, *i.e.*, to verify that the GMM estimator based on the whole set of "intuitive"' conditions is efficient, we need to compare the number of conditions obtained above to the number of conditions identified from covariance restrictions.

Under conditions (5) to (8) above, we can identify conditions on off-diagonal terms (corresponding to covariances between $u_{it}$ and $u_{is}$). Straightforward algebra reveals that we have the $(T-2)(T-3)/2$ conditions:

$$\frac{E[u_{it}(u_{i,s+1} - u_{is})]}{\theta_{s+1} - \theta_s} \text{ is the same } \forall t = 4, \ldots, T, \quad s = 1, \ldots, t-2.$$

Furthermore, there are additional conditions that can be found on covariances between the $u_{it}$'s and initial conditions $y_{i0}$. We have the following conditions:

$$\frac{E[y_{i0}(u_{it} - u_{i1})]}{\theta_t - \theta_1} \text{ is the same } \forall t = 2, \ldots, T,$$

which account for $T-2$ restrictions.

Note that $\triangle\theta_2 = \theta_2 - \theta_1$ is not identified in our model, in a similar fashion to the standard MTW case, where $\theta_1$ is not. If we set arbitrarily $\triangle\theta_2 = 1$, we have the following $T-3$ conditions:

$$E\left(\frac{(u_{i2} - u_{i1})(u_{i,t+1} - u_{it})}{(\theta_{t+1} - \theta_t)}\right) \qquad \text{is the same } \forall t = 3, \ldots, T.$$

In sum, the total number of moment conditions is $(T-2)(T-3)/2 + (T-2) + (T-3) = (T-1)(T-2)/2 + (T-3)$. Hence, the number of conditions obtained from covariance restrictions corresponds to the number of our "intuitive" moment conditions.

Finally, let us consider the homoskedasticity assumption for $\varepsilon_{it}$:

$$\text{for all } i, \quad E(\varepsilon_{it}^2) \text{ is the same } \forall t = 1,\ldots,T.$$

Tedious but simple algebra reveals that we have the moment conditions:

$$E\left(\frac{(\sum_{t=1}^{T} u_{it})(u_{it} - u_{is})}{\theta_t - \theta_s}\right) \quad \text{is the same } \forall t \neq s,$$

which yields $T - 2$ additional conditions.

# 4 The model with explanatory variables

We now consider the augmented model

$$\begin{aligned} y_{it} &= \delta y_{i,t-1} + \beta x_{it} + u_{it}, \\ u_{it} &= \alpha_i + \theta_t v_i + \varepsilon_{it}, \end{aligned} \tag{15}$$

$i = 1,2,...,N$, $t = 1,2,...,T$, where $x_{it}$ is a $K \times 1$ vector of time-varying explanatory variables, and $\beta$ is the $1 \times K$ vector of associated parameters.

As in the case of LOW and MTW models, when explanatory variables are introduced in the model, supplementary moment conditions may be considered in addition to those derived above, and are indeed often necessary for identifying the augmented set of parameters. These moment conditions will depend upon assumptions imposed on the $x_{it}$'s along different correlation patterns: a) correlation between the $x_{it}$'s and the $\varepsilon_{it}$'s; b) correlation between the $x_{it}$'s and $\alpha_i$; c) correlation between the $x_{it}$'s and $v_i$.

Let us for now assume that $K = 1$, and review the different possible cases.

*Case 1.a.*

$x_{it}$ is strictly exogenous, *i.e.*,

$$E(x_{is}\varepsilon_{it}) = 0 \quad \text{for} \quad s = 1,\ldots,T, \ t = 1,\ldots,T,$$

and correlated with both $\alpha_i$ and $v_i$.

In this case we have the following $T(T-2)$ moment conditions:

$$E[x_{is}(\triangle u_{it} - \tilde{r}_t \triangle u_{i,t-1})] = 0, \qquad t = 3,\ldots,T, \qquad s = 1,\ldots,T. \tag{16}$$

No extra restrictions are available from the equation in level, first difference or quasi difference.

*Case 1.b.*

$x_{it}$ is strictly exogenous, correlated with $\alpha_i$ but not correlated with $v_i$.

Then, in addition to (16), there are available moment conditions using $x_{is}$ $(s = 1,\ldots,T)$ as instruments in the equation in first difference:

$$(17) \qquad E(x_{is}\triangle u_{iT}) = 0 \quad s = 1,\ldots,T.$$

We have only $T$ extra restrictions, as the remaining ones have already been exploited for the equation in double difference.

*Case 1.c.*

$x_{it}$ is strictly exogenous, correlated with $v_i$ but not correlated with $\alpha_i$.

In this case, moment conditions from the model in quasi difference are not redundant with the set of conditions (16), unless $r_t = 1\ \forall t$. This is because moment restrictions (16) can be expressed as combinations of restrictions from the model in first difference, but not from the model in quasi difference when $r_t \neq 1$. Hence, we have the $T(T-1)$ extra moment conditions:

$$(18) \qquad E[x_{is}(u_{it} - r_t u_{i,t-1})] = 0 \quad s = 1,\ldots,T, \quad t = 2,\ldots,T.$$

*Case 1.d.*

$x_{it}$ is strictly exogenous and neither correlated with $\alpha_i$ nor with $v_i$.

These assumptions provide the greater number of conditions, because the $x_{is}$'s, $(s = 1,\ldots,T)$ can now be used as instruments for the equation in level. We have $2T$ extra restrictions:

$$E(x_{is}u_{it}) = 0 \quad s = 1,\ldots,T, \quad t = T-1,T.$$

Note that for the LOW and MTW models, exploiting the restrictions of no-correlation with the individual effects would provide us with less extra moment conditions. We would have the $T$ extra restrictions:

$$E(x_{is}u_{iT}) = 0, \quad s = 1,\ldots,T,$$

in addition to $E(x_{is}\triangle u_{it}) = 0$ for model LOW and $E[x_{is}(u_{it} - r_t u_{i,t-1})] = 0$ for MTW, $t = 2,\ldots,T$, $s = 1,\ldots,T$.

*Case 2.a.*

$x_{it}$ is weakly exogenous or predetermined, *i.e.*,

$$(19) \qquad \begin{array}{ll} E(x_{is}\varepsilon_{it}) = 0 & \text{for} \quad s = 1,\ldots,t, \quad t = 1,\ldots,T, \\ E(x_{is}\varepsilon_{it}) \neq 0 & \text{for} \quad s = t+1,\ldots,T, \quad t = 1,\ldots,T \end{array}$$

and correlated with $\alpha_i$ and $v_i$.

In this case, we have the following $(T-1)(T-2)/2$ moment conditions:

$$(20) \quad E[x_{is}(\triangle u_{it} - \tilde{r}_t\triangle u_{i,t-1})] = 0, \qquad s = 1,\ldots,t-2, \qquad t = 3,\ldots,T.$$

As in case 1.a., we cannot exploit extra restrictions from the equation in level, first difference or quasi difference. Imposing different correlation patterns between the $x_{it}$'s and the individual effects will however supply us with the same type of conditions as above.

*Case 2.b.*

$x_{it}$ is weakly exogenous and correlated with $\alpha_i$ but not correlated with $v_i$.

Then, there are $T-1$ extra moment conditions in addition to (20):

$$(21) \qquad E(x_{i,s-1}\triangle u_{is}) = 0 \quad s = 2,\ldots,T.$$

*Case 2.c.*

$x_{it}$ is weakly exogenous and correlated with $v_i$ but not correlated with $\alpha_i$.

We have $T(T-1)/2$ additional moment restrictions:

$$(22) \qquad E[x_{is}(u_{it} - r_t u_{i,t-1})] = 0 \quad s = 1,\ldots,t-1, \quad t = 2,\ldots,T.$$

*Case 2.d.*

$x_{it}$ is weakly exogenous and neither correlated with $\alpha_i$ nor with $v_i$.

In this case where moment conditions from the equation in level can be exploited, we have $2T - 1$ extra restrictions:

$$E(x_{i1}u_{i1}) = 0 \ \text{ and } \ E(x_{is}u_{it}) = 0 \ \ s = t-1, t, \quad t = 2,\ldots,T.$$

Finally, note that, as above, for the LOW and MTW models, the assumption of no-correlation with the individual effects would yield only $T$ extra moment conditions:

$$E(x_{is}u_{is}) = 0, \quad s = 1,\ldots,T,$$

in addition to $E(x_{is}\triangle u_{it}) = 0$ for model LOW and $E[x_{is}(u_{it} - r_t u_{i,t-1})] = 0$, for MTW, $t = 2,\ldots,T$, $s = 1,\ldots,t-1$.

Quite obviously, if there is more than one explanatory variable, the same reasoning will apply. Assumptions regarding each variable will determine the corresponding additional available moment conditions.

We have assumed so far that the explanatory variables were time-varying. To conclude, it is worth considering the special situation where one or more of the explanatory variables are time-invariant. Let us assume for now that there is a single time-invariant explanatory variable $z_i$.[8] Identification of the parameter associated with $z_i$ crucially depends on correlation assumptions with the individual effects, and may not always be achieved.

We need to assume that $z_i$ is exogenous:[9]

$$E(z_i\varepsilon_{it}) = 0 \ \forall t,$$

and that $z_i$ is not correlated with $\alpha_i$ to get additional moment conditions available, whatever the correlation pattern between $z_i$ and $v_i$. Note, however, that the parameter associated with $z_i$ may not be identified. This would be the case if no other moment conditions (associated with other explanatory variables) implying the equation in level or in quasi difference were available. The fact that the model contains or not extra explanatory variables is therefore crucial.

*Case a*

$z_i$ is correlated with $v_i$. We get the following moment condition:

$$(23) \qquad E\left[z_i \frac{1}{T} \textstyle\sum_{t=2}^{T}(u_{it} - r_t u_{i,t-1})\right] = 0.$$

8. The same reasoning applies to each time-invariant variable if $z_i$ contains more than one component.

9. The distinction between strict and weak exogeneity is not meaningful here.

*Case b*
$z_i$ is neither correlated with $\alpha_i$ nor with $v_i$. In this case we get, in addition to (23), the extra moment condition:

(24) $$E\left(z_i \frac{1}{T} \sum_{t=1}^{T} u_{it}\right) = 0.$$

# 5 Monte Carlo experiments

We investigate in this section the behavior of our double-difference estimator, as well as the alternative first-difference and quasi-difference estimators, for dynamic panel data models, using Monte Carlo techniques. The main purpose of the simulation experiment is to evaluate the small sample bias of the GMM estimators and the actual size of their associated Hansen specification test, when the true data generating process is specification (4), LOW, or MTW.

We first consider the dynamic model with a data generating process corresponding to (4). Throughout the experiment, $N$ is set to 200 and $T$ is 6. In all simulations, we set $\delta = 0.7$ and $\sigma_\varepsilon = 0.2$. The number of replications is 10,000.

We generate individual effects from a unique distribution, assuming that the variance of $\alpha_i$ and $v_i$ is the same for all individuals. For different sets of parameters $(\sigma_\alpha, \sigma_v, \rho_{\alpha,v})$, $v_i$ and $\varepsilon_{it}$ are generated from normal distributions with mean 0 and variances $\sigma_v^2$, $\sigma_\varepsilon^2$ respectively. To allow for correlation between effects $\alpha_i$ and $v_i$, we generate $\alpha_i$ from its conditional distribution given $v_i$.

Initial values of the process are then drawn conditionally upon $\alpha_i$ under the stationarity assumption:

$$y_{i0} \sim N\left(\frac{\alpha_i + \theta_0 v_i}{1-\delta}, \frac{\sigma_\varepsilon^2}{1-\delta^2}\right).$$

We finally obtain $N \times T$ simulated draws for $y$ as

$$y_{it} = \delta y_{i,t-1} + \alpha_i + \theta_t v_i + \varepsilon_{it}.$$

For time effects $\theta_t$, two specifications are chosen. First, we set $\theta_t$: $\theta_t = \exp(-t/2)$. Second, to explore the case where $\theta_t$ is not monotonic, we set $\theta_t = 0.5 + 0.2(t+2) - 0.02(t+2)^2$. With this specification, $\theta_t$ is a concave, bell-shaped function with a maximum around $t = 3$.

Six sets of parameters are used:

$(\sigma_\alpha = 1.0, \sigma_v = 1.0, \rho_{\alpha v} = 0.0)$; $(\sigma_\alpha = 1.0, \sigma_v = 1.0, \rho_{\alpha v} = 0.5)$;
$(\sigma_\alpha = 0.5, \sigma_v = 1.0, \rho_{\alpha v} = 0.0)$; $(\sigma_\alpha = 0.5, \sigma_v = 1.0, \rho_{\alpha v} = 0.5)$;
$(\sigma_\alpha = 1.0, \sigma_v = 0.5, \rho_{\alpha v} = 0.0)$ and $(\sigma_\alpha = 1.0, \sigma_v = 0.5, \rho_{\alpha v} = 0.5)$.

To save space and computer time, we restrict ourselves to estimators using the Arellano and Bond instruments for all cases, as Ahn-Schmidt instruments would not be valid when the lower set of instruments used for the ARELLANO-BOND like estimator is not.

For each of the two processes considered for parameters $\theta_t$, we report the mean and standard error of GMM estimates obtained from the first-, quasi- and double-difference estimators, as well as the 1 %, 5 % and 10 % sizes of their corresponding Hansen test statistic. When the true data generating process is our mixed structure, the proportion of rejections of the null hypothesis for our double-difference GMM estimator is the small-sample size of Hansen test, and should be close to corresponding critical values (0.01, 0.05 and 0.1). Moreover, for GMM estimators based on either first difference or quasi difference, the proportion of rejections is the small-sample power of Hansen test.

Table 1 presents simulation results for the negative exponential process. The bias in the first-difference GMM estimator is higher than the bias in our double-difference estimator, for all sets of parameters $\sigma_\alpha$, $\sigma_v$ and $\rho_{\alpha v}$, while the standard error is lower. In this case, the negative exponential form is in fact close to a linear representation, and as a consequence, first-difference transformation succeeds in filtering out a higher proportion of the full unobserved heterogeneity term $\alpha_i + \theta_t v_i$. Therefore, the bias in the first-difference GMM estimator is smaller when the variance of the additive individual effect $\alpha_i$ is larger than the variance of the multiplicative effect $v_i$. Note that the first-difference GMM tends to underestimate the true value of parameter $\delta$, for all parameter sets. Also, the power of the HANSEN test is lower when the ratio $\sigma_\alpha/\sigma_v$ is greater than 1.

The bias associated with the quasi-difference estimator is much larger than the one for the two other transformations. This estimator also has a very high proportion of rejections, which tends to be larger when $\sigma_\alpha/\sigma_v = 1$. The argument exposed above can be inverted here: because the quasi-difference transformation filters out the second effect $v_i$, the bias is higher when the ratio $\sigma_\alpha/\sigma_v$ increases, in particular, when $\sigma_\alpha > \sigma_v$. Also, the proportion of rejections is higher (for cases 3 to 6).

Concerning our double-difference estimator, there remains a small-sample downward bias in all cases, between -0.0092 in case 2 to -0.0800 in case 6. The sizes of Hansen test are lower than corresponding critical values for cases 1 and 2. On the other hand, for cases 3 and 4, the model specification is over-rejected by the Hansen test. In cases 5 and 6, when $\sigma_\alpha = \sigma_v$, the proportion of rejections is very close to critical values.

Simulation results for the bell-shape process are presented in Table 2. Because this specification entails a non-monotonic process for $\theta_t$, the first-difference estimator is expected to perform worse than in the previous case. As can be seen, for the first-difference estimator, although standard errors are always lower, the bias and the power of Hansen test are more important than before, in general. The bias of the first-difference estimator is the smallest in cases 1 and 2, where $\sigma_\alpha/\sigma_v = 2$. Again, when $\alpha$ has a greater contribution to total unobserved heterogeneity, first-differencing the model eliminates a greater proportion of the full individual heterogeneity. However, the relative magnitude of the variances does not seem to have an impact on the behavior of the quasi-difference estimator for this specification of $\theta_t$. This estimator

TABLE 1
***Small-sample properties of GMM estimators***
***Data Generating Process:*** $u_{it} = \alpha_i + \theta_t v_i + \varepsilon_{it} \quad \theta_t = e^{-t/2}$

| Procedure | Mean | Std err | 1 % | 5 % | 10 % |
|---|---|---|---|---|---|
| Case 1. $\sigma_\alpha = 1, \sigma_v = 0.5, \rho_{\alpha v} = 0.5$ | | | | | |
| FD | 0.6399 | 0.1235 | 0.0147 | 0.0915 | 0.2108 |
| QD | 0.8590 | 0.1355 | 0.2759 | 0.5002 | 0.6354 |
| DD | 0.6660 | 0.3201 | 0.0048 | 0.0228 | 0.0537 |
| Case 2. $\sigma_\alpha = 1, \sigma_v = 0.5, \rho_{\alpha v} = 0.0$ | | | | | |
| FD | 0.6537 | 0.1490 | 0.0174 | 0.1048 | 0.2319 |
| QD | 0.9645 | 0.1705 | 0.3224 | 0.5478 | 0.6826 |
| DD | 0.6908 | 0.3067 | 0.0059 | 0.0276 | 0.0584 |
| Case 3. $\sigma_\alpha = 0.5, \sigma_v = 1, \rho_{\alpha v} = 0.5$ | | | | | |
| FD | 0.5033 | 0.1987 | 0.2017 | 0.5004 | 0.6898 |
| QD | 0.9669 | 0.2670 | 0.5369 | 0.6588 | 0.7939 |
| DD | 0.6409 | 0.3877 | 0.0233 | 0.0838 | 0.1436 |
| Case 4. $\sigma_\alpha = 0.5, \sigma_v = 1, \rho_{\alpha v} = 0.0$ | | | | | |
| FD | 0.5056 | 0.2328 | 0.3830 | 0.6544 | 0.7788 |
| QD | 0.9407 | 0.2387 | 0.4364 | 0.7438 | 0.8703 |
| DD | 0.6338 | 0.3983 | 0.0218 | 0.0783 | 0.1330 |
| Case 5. $\sigma_\alpha = 1, \sigma_v = 1, \rho_{\alpha v} = 0.0$ | | | | | |
| FD | 0.6083 | 0.1385 | 0.1982 | 0.4882 | 0.6771 |
| QD | 0.9540 | 0.2144 | 0.9769 | 0.9815 | 0.9850 |
| DD | 0.6761 | 0.3127 | 0.0105 | 0.0589 | 0.1058 |
| Case 6. $\sigma_\alpha = 1, \sigma_v = 1, \rho_{\alpha v} = 0.5$ | | | | | |
| FD | 0.5970 | 0.1473 | 0.1143 | 0.3555 | 0.5468 |
| QD | 0.9727 | 0.1609 | 0.7935 | 0.9060 | 0.9441 |
| DD | 0.6200 | 0.3015 | 0.0131 | 0.0565 | 0.1013 |

*Estimation procedure: Two-step GMM with Arellano-Bond instruments.*
*FD, QD and DD stand for first-difference, quasi-difference and double-difference transformations respectively.*
*1 %, 5 % and 10 % denotes respectively the 0.01, 0.05 and 0.1 size (power) of Hansen test for DD (FD, QD).*
*N is 200, T is 6. Based on 10,000 replications..*

has again a significant upward bias. Compared to the negative exponential process, the proportion of rejections of the Hansen test is higher for cases 1 to 4, but is lower for cases 5 and 6. Concerning our double-difference estimator, the small-sample bias is higher than for the negative exponential process in cases 3 to 5. Our procedure tends to overestimate the true value of $\delta$, and the smallest bias is for case 2, as before. The Hansen test now over-rejects our model specification except for case 2, where test sizes are lower than critical values.

TABLE 2
***Small-sample properties of GMM estimators***
***Data Generating Process:*** $u_{it} = \alpha_i + \theta_t v_i + \varepsilon_{it}$
$\theta_t = 0.5 - 0.02(t+2)^2 + 0.2(t+2)$

| Procedure | Mean | Std err | 1 % | 5 % | 10 % |
|---|---|---|---|---|---|
| Case 1. $\sigma_\alpha = 1, \sigma_v = 0.5, \rho_{\alpha v} = 0.5$ | | | | | |
| FD | 0.6355 | 0.0528 | 0.3767 | 0.7041 | 0.8440 |
| QD | 0.8805 | 0.0733 | 0.7217 | 0.8691 | 0.9235 |
| DD | 0.7492 | 0.1533 | 0.0207 | 0.0792 | 0.1450 |
| Case 2. $\sigma_\alpha = 1, \sigma_v = 0.5, \rho_{\alpha v} = 0.0$ | | | | | |
| FD | 0.6570 | 0.0515 | 0.0916 | 0.2938 | 0.4823 |
| QD | 0.9119 | 0.0775 | 0.6733 | 0.8408 | 0.9071 |
| DD | 0.7007 | 0.1419 | 0.0093 | 0.0459 | 0.0880 |
| Case 3. $\sigma_\alpha = 0.5, \sigma_v = 1, \rho_{\alpha v} = 0.5$ | | | | | |
| FD | 0.5382 | 0.0414 | 0.9801 | 0.9855 | 0.9914 |
| QD | 0.8791 | 0.0750 | 0.7103 | 0.8652 | 0.9214 |
| DD | 0.8404 | 0.1417 | 0.0328 | 0.1238 | 0.2115 |
| Case 4. $\sigma_\alpha = 0.5, \sigma_v = 1, \rho_{\alpha v} = 0.0$ | | | | | |
| FD | 0.5252 | 0.0517 | 1.0000 | 1.0000 | 1.0000 |
| QD | 0.9342 | 0.0525 | 0.8040 | 0.8469 | 0.9093 |
| DD | 0.8290 | 0.1451 | 0.0373 | 0.1282 | 0.2147 |
| Case 5. $\sigma_\alpha = 1, \sigma_v = 1, \rho_{\alpha v} = 0.5$ | | | | | |
| FD | 0.5920 | 0.0511 | 0.9989 | 1.0000 | 1.0000 |
| QD | 0.9321 | 0.0536 | 0.6656 | 0.8400 | 0.9054 |
| DD | 0.8163 | 0.1644 | 0.0331 | 0.1187 | 0.2007 |
| Case 6. $\sigma_\alpha = 1, \sigma_v = 1, \rho_{\alpha v} = 0.0$ | | | | | |
| FD | 0.6030 | 0.0514 | 0.9722 | 0.9965 | 0.9990 |
| QD | 0.9020 | 0.0772 | 0.6012 | 0.7956 | 0.8757 |
| DD | 0.7871 | 0.1712 | 0.0333 | 0.1184 | 0.2056 |

*Estimation procedure: Two-step GMM with Arellano-Bond instruments.*
*FD, QD and DD stand for first-difference, quasi-difference and double-difference transformations respectively.*
*1 %, 5 % and 10 % denotes respectively the 0.01, 0.05 and 0.1 size (power) of Hansen test for DD (FD, QD).*
*N is 200, T is 6. Based on 10,000 replications.*

To investigate the behavior of our estimator under alternative data generating processes, we conduct a second simulation experiment where errors are generated according to LOW and MTW schemes. Table 3 presents the mean and standard error of GMM estimates, together with the proportion of rejections of Hansen test statistic for the three alternative estimators (first difference for LOW, quasi difference for MTW and double difference for model (4)).

TABLE 3
***Small-sample behavior of GMM estimators***
***Data generating process:***
***LOW:*** $u_{it} = \alpha_i + \varepsilon_{it}$, ***MTW:*** $u_{it} = \theta_t v_i + \varepsilon_{it}$.

| Process | Mean | Std err | 1 % | 5 % | 10 % |
|---|---|---|---|---|---|
| Case A. True process: LOW | | | | | |
| FD estimate | 0.6205 | 0.1559 | 0.0139 | 0.0649 | 0.1301 |
| QD estimate | 0.7779 | 0.2015 | 0.0190 | 0.0802 | 0.1466 |
| DD estimate | 0.6173 | 0.3813 | 0.0055 | 0.0232 | 0.0538 |
| Case B. True process: MTW, negative exponential | | | | | |
| FD estimate | 0.7768 | 0.0919 | 0.1595 | 0.3810 | 0.5222 |
| QD estimate | 0.6822 | 0.0774 | 0.0048 | 0.0343 | 0.0769 |
| DD estimate | 0.7582 | 0.2941 | 0.0098 | 0.0421 | 0.0850 |
| Case C. True process: MTW, bell-shape | | | | | |
| FD estimate | 0.6160 | 0.0525 | 0.1660 | 0.5101 | 0.7942 |
| QD estimate | 0.7051 | 0.0434 | 0.0122 | 0.0628 | 0.1255 |
| DD estimate | 0.7551 | 0.0992 | 0.0388 | 0.1335 | 0.2215 |

*Estimation procedure: Two-step GMM with Arellano-Bond instruments.*
*FD, QD and DD stand for first-difference, quasi-difference and double-difference transformations respectively.*
*For LOW,* $\sigma_\alpha = 1.0$. *For MTW,* $\sigma_v = 1.0$.
*Exponential case for MTW:* $\theta_t = \exp(-t/2)$.
*Bell-shape case for MTW:* $\theta_t = 0.5 - 0.02(t+2)^2 + 0.2(t+2)$.
*1 %, 5 % and 10 % denotes respectively the 0.01, 0.05 and 0.1 size of Hansen test for FD, QD and DD.*
*N is 200, T is 6. Based on 10,000 replications.*

We can see that in all situations, our estimator has a higher small-sample bias and is less efficient than the relevant estimator, as expected. The standard error of the double-difference estimator is about twice as much as the alternative estimator, except when $\theta$ is negative exponential, where it is about four times as much as the quasi-difference one. Interestingly, the quasi-difference estimator performs relatively well when model LOW is the true data generating process (case A). Its bias is close to the first-difference estimator one, but it is less efficient. Note also that the proportion of rejections is much dependent on the data generating process. When the true process is LOW or MTW-negative exponential, the sizes of Hansen test for our estimator are below critical values. On the other hand, the validity of moment conditions is rejected more often in the third case (MTW and bell-shape). Test sizes associated with usual (first-difference or quasi-difference) estimators are slightly higher than critical values in cases A and C, but lower than critical values for case B.

# 6 Concluding remarks: an empirical strategy for estimation and testing

This paper introduced a new error-component structure for dynamic models with panel data, that contains as special cases two standard specifications found in the literature. Our error specification consists of a single additive individual effect and a time-varying individual effect, both being possibly correlated. We show that, under some conditions on the variance-covariances of individual effects and on time-varying parameters, GMM estimators based on first difference or quasi difference and using lagged dependent variables as instruments are not consistent. Even under the assumption of no serial correlation of error terms, first-difference or quasi-difference transformation do not eliminate both individual effects. As a consequence, usual procedures do not lead to valid moment conditions.

We present a simple double-difference transformation that filters out both components in the unobserved heterogeneity term, and we show how to retrieve the maximum number of moment conditions for GMM estimation. These moment conditions are expressed in a manner similar to AHN and SCHMIDT [1995], and match conditions based on covariance restrictions.

The small-sample behavior of GMM estimators applied to our model specification is investigated with a Monte Carlo experiment. For our double-difference estimator, the size of Hansen test is close to critical values when $\theta$ is negative exponential, but often higher than critical values when $\theta$ is bell-shaped. The power of Hansen test (when assuming moment conditions derived from LOW – Linear One-Way – or MTW – Multiplicative Two-Way – model specification) is highly dependent on the ratio of individual effect variances, in particular for the first-difference estimator. As far as the quasi-difference estimator is concerned, the power of Hansen test is much more important in all cases, especially when $\theta$ is bell-shaped.

Simulation results also show that GMM with double-differencing results in a general and significant loss in efficiency when the true data generating process is the LOW or the MTW model. This poor efficiency is on one hand due to the fact that the double-difference transformation reduces the number of time periods that may be used, resulting in a lower number of moment conditions. Compared to the GMM estimator based on the first-difference transformation, we have less active moment conditions ($(T-1)(T-2)/2+(T-3)$ instead of $T(T-1)/2+(T-2)$) when the matrix of instruments consists of lagged dependent variables only.[10] The same is true for MTW, with the same number of moment conditions as for the first-difference GMM estimator. This lack of efficiency is on the other hand also due to the reduction in the variability of the data, and hence in the quantity of information contained in the observations, when a double-difference transformation is used, compared to a single-difference transformation. To conclude on our simulation experiments, the poor efficiency of the GMM estimator with double-differencing, when LOW or

10. That is, we have $T$ additional conditions when using first difference instead of double difference.

MTW is the true data generating process, can be constrasted with the far more satisfactory behavior of the GMM estimator using first difference when the true data generating process is the LOW model (Table 3, Case A).

The above discussion on efficiency has important consequences for the estimation and testing strategy. On the one hand, as the true data generating process (either LOW, MTW or specification (4)) is not known beforehand, it would be tempting to consider our double-difference GMM as an "all-purpose" estimator. Indeed, LOW and MTW models are embedded in our specification, resulting in a double-difference GMM estimator which is always consistent under these alternative specifications.[11] On the other hand, as detailed above and confirmed by our simulation results, there is a significant reduction in efficiency when our GMM estimator is used while the true model is LOW or MTW. The same can be said about the strategy consisting in using the quasi-difference GMM estimator as an "all-purpose" estimator, perhaps not because of efficiency loss, but primarily because this estimator would not be consistent under specification (4).

Therefore, it is important from an empirical point of view to identify the genuine model specification. To do so, we suggest the three following strategies, each based on a series of estimation and testing steps.

*Strategy 1*

We start here with the first-difference GMM estimator, which is more efficient when the LOW specification is true, and test for the validity of moment conditions using the Hansen test statistic. If model specification is rejected, a specification check should be performed under alternative serial correlation patterns. This will typically amount to excluding particular lagged dependent variables from the original Arellano-Bond matrix of instruments, see, *e.g.*, ARELLANO and BOND [1991]. If, even after accounting for serial correlation in the $\varepsilon_{it}$'s, the LOW model specification is rejected, one should consider the MTW model specification next. As before, if the Hansen test rejects the null, checking for possible serial correlation in the $\varepsilon_{it}$'s is recommended. Finally, in the case where both LOW and MTW models, including their serial-correlation versions, are rejected by Hansen test statistics, our estimator based on a double-difference transformation can be computed.

*Strategy 2*

With this intermediate strategy, we would first estimate the model by GMM with quasi-difference, and test for the validity of the associated moment conditions. If the associated Hansen test rejects the null, we use our double-difference GMM estimator and perform another Hansen specification test. If, on the other hand, the Hansen test for the validity of moment conditions associated with the MTW model does not reject the null, we pursue by testing for the LOW specification, using a Wald test based on quasi-difference GMM estimates (the Wald test here would check for the constancy of time parameters $r_t$). If the Wald test does not reject the null, we then estimate the model by first-difference GMM and perform a Hansen specification test for the validity of the moment conditions associated with the LOW model.

11. Note that this line of reasoning would also be in favor of using, in one-way panel data models, the quasi-difference GMM estimator, which is consistent when model LOW is the true data generating process.

*Strategy 3*

This is a "descending" strategy, by which we consider the consistent but not always efficient GMM estimator first. In this case, we start with the double-difference GMM estimator and test for the validity of moment conditions with the Hansen test. If the latter does not reject the null, we test for moment conditions associated with model MTW (*i.e.*, by testing for the validity of those non-redundant moments corresponding to the MTW specification, based on double-difference GMM estimates, see HALL [1991]). If the null is not rejected, we estimate the model by quasi-difference GMM and perform a Hansen test. If moment conditions are not rejected, we test for the validity of the error structure implied by the LOW model using a Wald test, and estimate the model by first-difference GMM if the model specification is not rejected. Finally, we perform a Hansen test based on first-difference GMM estimates.

It is important to note that, whatever the strategy adopted, rejecting specification LOW, MTW or (4) by the Hansen test does not necessarily mean that the error structure is misspecified. It may well be the case that the assumption of no-serial correlation in the $\varepsilon_{it}$'s, as introduced in Section 2, is not verified by the data. $\varepsilon_{it}$ may follow a $MA(q)$ process, $q \geqslant 1$, or worse, an $AR(1)$ process, in which case lagged dependent variables are never valid instruments, and model parameters are only identified through the introduction of explanatory variables. When the model specification is rejected, it is therefore necessary to perform other specification checks (test for serial correlation in the $\varepsilon_{it}$'s, but also checks for other sources of misspecification such as linearity tests in the conditional mean).

The advantages and drawbacks associated with the empirical estimation and testing strategies suggested above are the following. First, with Strategy 1, we clearly favor efficiency by looking for the first available consistent GMM estimator when considering less and less restrictive moment conditions. Efficiency is an important issue for dynamic panel data estimation, where the presence of weak instruments is frequent in practice. Also, by starting with the first-difference GMM estimator, we do not have to use nonlinear optimization techniques to compute an estimator which is consistent under the assumption that the LOW model is true. Second, Strategy 3 may seem more intuitive at first, because our double-difference GMM estimator is consistent under alternative specifications (LOW or MTW), which obtain as special cases of model (4). We nevertheless feel that, for theoretical reasons supported by our simulation results, the loss in efficiency entailed by Strategy 3 may be too severe when the LOW or the MTW model is the true data generating process. Finally, Strategy 2 beginning with the quasi-difference GMM estimator has the obvious disadvantage of an intermediate strategy: this strategy requires a nonlinear minimization algorithm from the start, it may be less efficient than the first-difference GMM estimator when model LOW is true, and is not consistent when specification (4) is the actual data generating process. For all these reasons, the recommended strategy would be Strategy 1.

# 7 Acknowledgments

We thank Manuel ARELLANO, Badi BALTAGI, Richard BLUNDELL, Jean-Pierre FLORENS, Guido IMBENS, Pascal LAVERGNE and Patrick SEVESTRE for helpful comments. Comments from the editor and an anonymous referee are gratefully acknowledged. ▼

# • References


AHN S.C., HOON LEE Y., SCHMIDT P. (2001). – "GMM Estimation of Linear Panel Data Models with Time-Varying Individual Effects", *Journal of Econometrics*, 101, p. 219-255.

AHN S.C., SCHMIDT P. (1995). – "Efficient Estimation of Models for Dynamic Panel Data", *Journal of Econometrics*, 68, p. 5-27.

AHN S.C., SCHMIDT P. (1995). – "Separability Result for GMM Estimation, With Applications to GLS Prediction and Conditional Moment Tests", *Econometric Reviews*, 14, 1, p. 19-34.

AHN S.C., SCHMIDT P. (1997). – "Efficient Estimation of Dynamic Panel Data Models: Alternative Assumptions and Simplified Estimation", *Journal of Econometrics*, 76, p. 309-321.

ALVAREZ J., ARELLANO M. (1998). – "The Time Series and Cross Section Asymptotics of Dynamic Panel Data Estimators", Working Paper n° 9808, CEMFI.

AMEMIYA T., MACURDY T.E. (1986). – "Instrumental-Variable Estimation of an Error-Components Model", *Econometrica*, 54, 4, p. 869-880.

ANDERSON T.W., HSIAO C. (1982). – "Formulation and Estimation of Dynamic Models Using Panel Data", *Journal of Econometrics*, 18, p. 47-82.

ARELLANO M., BOND S. (1991). –"Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58, p. 277-297.

ARELLANO M., BOVER O. (1995). – "Another Look at the Instrumental Variable Estimation of Error-Components Models", *Journal of Econometrics*, 68, p. 29-51.

BLUNDELL R., BOND S. (1998). – "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics*, 87, p. 115-143.

BREUSCH T.S., MIZON G.E., SCHMIDT P. (1989). – "Efficient Estimation Using Panel Data", *Econometrica*, 57, 3, p. 695-700.

CHAMBERLAIN G. (1987). – "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", *Journal of Econometrics*, 34, p. 305-334.

CHAMBERLAIN G. (1987). – "Comment: Sequential Moment Restrictions in Panel Data", *Journal of Business and Economic Statistics*, 10, 20-26.

CRÉPON B., KRAMARZ F., TROGNON A. (1997). – "Parameters of Interest, Nuisance Parameters and Orthogonality Conditions An Application to Autoregressive Error Component Models", *Journal of Econometrics*, 82, p. 135-156.

HALL A.R (1999). – "Hypothesis testing in models estimated by GMM", in L. Matyas (ed.), *Generalized Method of Moments Estimation*, Cambridge University Press, p. 96-127.

HAUSMAN J.A. (1978). – "Specification Tests in Econometrics", *Econometrica*, 46, 6, p. 1251-1271.

HAUSMAN J.A., TAYLOR W.E. (1981). – "Panel Data and Unobservable Individual Effects", *Econometrica*, 49, 6, p. 1377-1398.

HOLTZ-EAKIN D., NEWEY W., ROSEN H. (1988). – "Estimating Vector Autoregressions with Panel Data", *Econometrica*, 56, p. 1371-1395.

HSIAO C., PESARAN M.H., TAHMISCIOGLU A.K. (1999). – "Bayes Estimation of Short-Run Coefficients in Dynamic Panel Data Models", in Hsiao C., K. Lahiri, L. Lee and M.H. Pesaran (eds.), *Analysis of Panels and Limited Dependent Variable Models*, Cambridge University Press, p. 268-296.

KIEFER N.M. (1980). – "A Time Series-Cross Section Model with Fixed Effects with an Intertemporal Factor Structure", unpublished manuscript, Cornell University.

LILLARD L., WEISS Y. (1979). – "Components of Variation in Panel Earnings Data: American Scientists 1960-1970", *Econometrica*, 47, p. 437-454.

MEGHIR C., WINDMEIJER F. (1999). – "Moment Conditions for Dynamic Panel Data Models with Multiplicative Individual Effects in the Conditional Variance", *Annales d'Économie et de Statistique*, 55-56, p. 317-330.

SCHMIDT P., AHN S. C., WYHOWSKI D. (1992). – "Comment: Sequential Moment Restrictions in Panel Data", *Journal of Business and Economic Statistics*, 10, p. 10-14.

## Books

BALTAGI B. H. (1995). – "Econometric Analysis of Panel Data", *John Wiley and Sons*.

DAVIDSON R., MACKINNON J.G. (1993). – "Estimation and Inference in Econometrics", *Oxford University Press*.

DORMONT B. (1989). – "Introduction à l'Économétrie des Données de Panel", *Éditions du Centre National de la Recherche Scientifique*.

HSIAO C. (1986). – "Analysis of Panel Data", *Cambridge University Press*.

MATYAS L., SEVESTRE P. (1992). – "The Econometrics of Panel Data Handbook of Theory and Applications", *Kluwer Academic Publishers*.